JAYAWANT SHIKSHAN PRASARAK MANDAL'S

Rajarshi Shahu College of Engineering

(An Autonomous Institute affiliated to SPPU)

Approved by AICTE, Accredited by NBA (UG Program)

Accredited by NAAC with "A" Grade, MHRD-NIRF Rank: 201-250

Engineering Design and Innovation

A.Y. 2021-22

ENVISAGING AND RETAINING COLLEGE WORKFORCE ATTRITION USING ML AND EL

The Team

Vasudev Surwase

Neha Patil

Janhvi Bhandekar

Jay Shinde

PROBLEM STATEMENT

To predict and retain college workforce attrition with various predictive models using Machine Learning and Ensemble Learning

INTRODUCTION

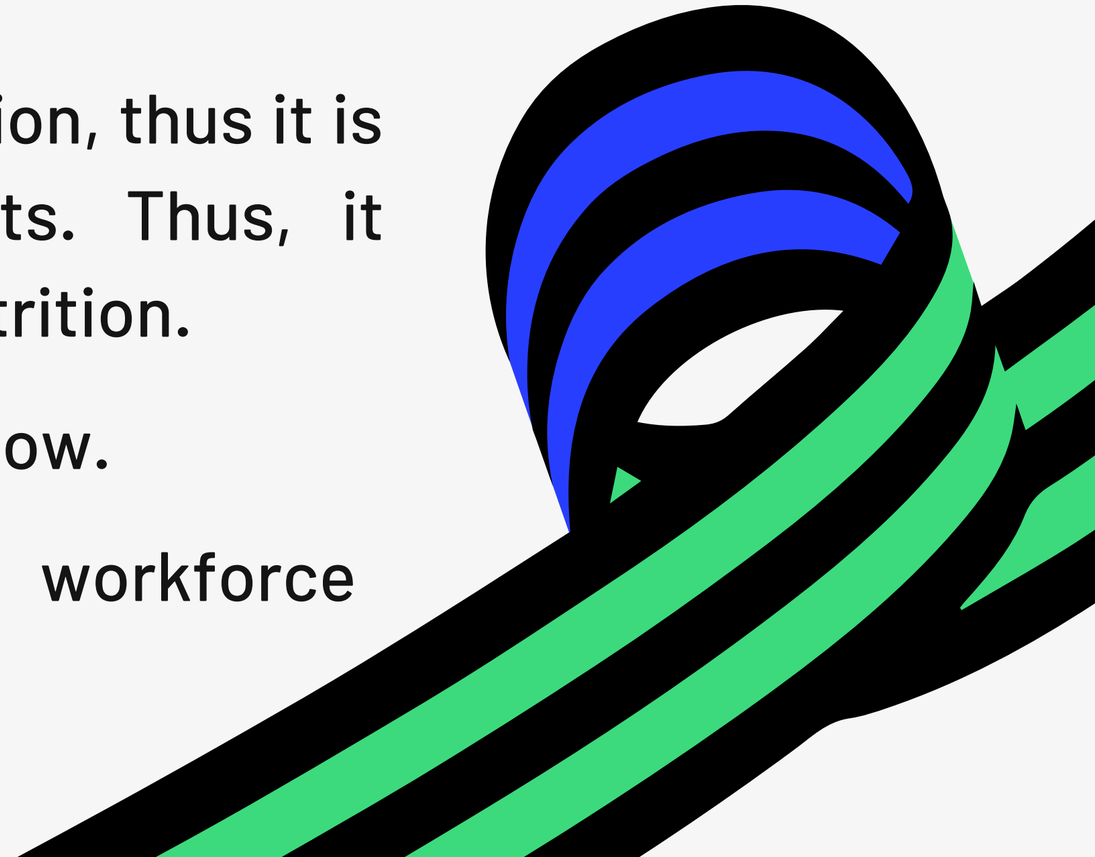
The process by which employees leave a company for any reason (voluntarily or unwillingly), of his own, is known as employee attrition.
In many industries, attrition is a serious issue that is currently at an all-time high.
Every organization must deal with this, which is one of the most important issues.
Employee attrition refers to the choice made by productive workers to quit the organization for many factors like work pressure, an unpleasant atmosphere, or inadequate pay.
The productivity of the organization is impacted by employee attrition.

OBJECTIVES

1. To apply preprocessing techniques as the data is collected from online survey, the data need to be cleaned and filtered.
2. To apply feature selection methods such as Recursive Feature Elimination (a wrapper method) and SelectKBest
3. Build decision models to predict attrition using Machine, Ensemble Learning techniques (ML, EL)
4. Make interpretations to explain and identify the exact reasons behind employee attrition
5. Make recommendations to fight this possible attrition and to take necessary HR management policies

MOTIVATION

- In the current environment, employee attrition is a big problem that requires attention.
- Employees are an asset to an organization, thus it is important to meet their requirements. Thus, it becomes crucial to predict employee attrition.
- No real time data set has been used till now.
- So we need to envisage the college workforce attrition using real time dataset.



TECHNOLOGIES

- **Frontend**

  HTML/ CSS/ JS
  MERN Stack

- **Backend**

  Python
  Flask
  Django

- **Database**

  MongoDB

HARDWARE REQUIREMENTS

- 8GM RAM
- i3 - i5 processor

Literature Survey

Research Paper	Year of Publishing	Title
IEEE	August 2021	A Novel Approach to Improve Employee Retention Using Machine Learning
IEEE	December 2021	Predicting Employee Attrition using Supervised Learning Classification Models
IEEE	December 2018	Predicting Employee Attrition using Machine Learning
IRJET	May 2020	Employee Attrition Predictive Model Using Machine Learning
IJEAT	October 2019	Antecedents to Employee Attrition Behavior in Indian IT/ITES Sector

OBSERVATION ON LITERATURE SURVEY

1. Machine Learning predictive algorithms namely, LR, DT, RT, SVM, Naive Bayes have been used largely.
2. Random Forest outperforms every algorithm in terms of accuracy.
3. Their is no method which will retain the employees based on their importance as particular employee plays particular role in an organization.
4. Only one dataset which is IBM Dataset of Kaggle is used in majority of papers, which might give irregular results for other organizations. Hence, the algorithm is not practised on real time dataset.
5. Feature Selection need to be done, to acquire the most useful attributes.

OBSERVATION ON LITERATURE SURVEY

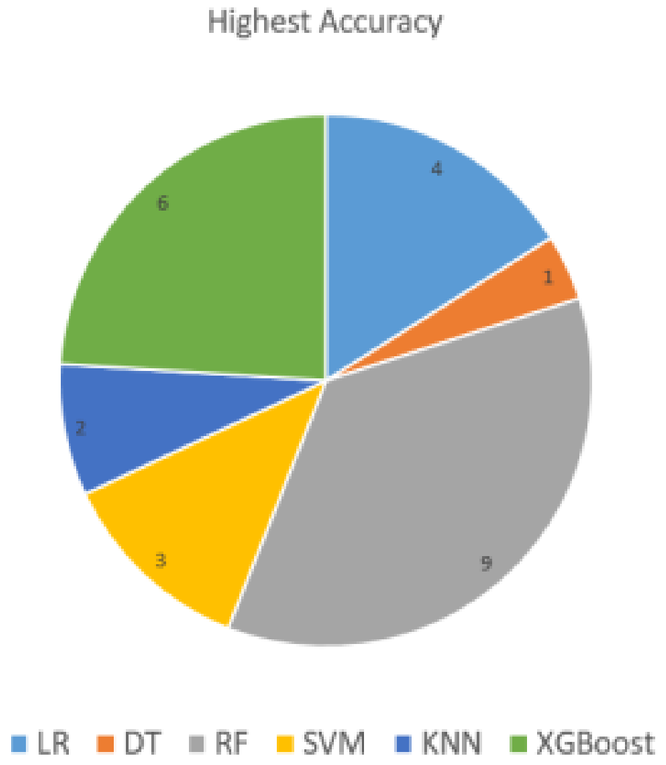


Fig 1: Count of algorithm with highest accuracy

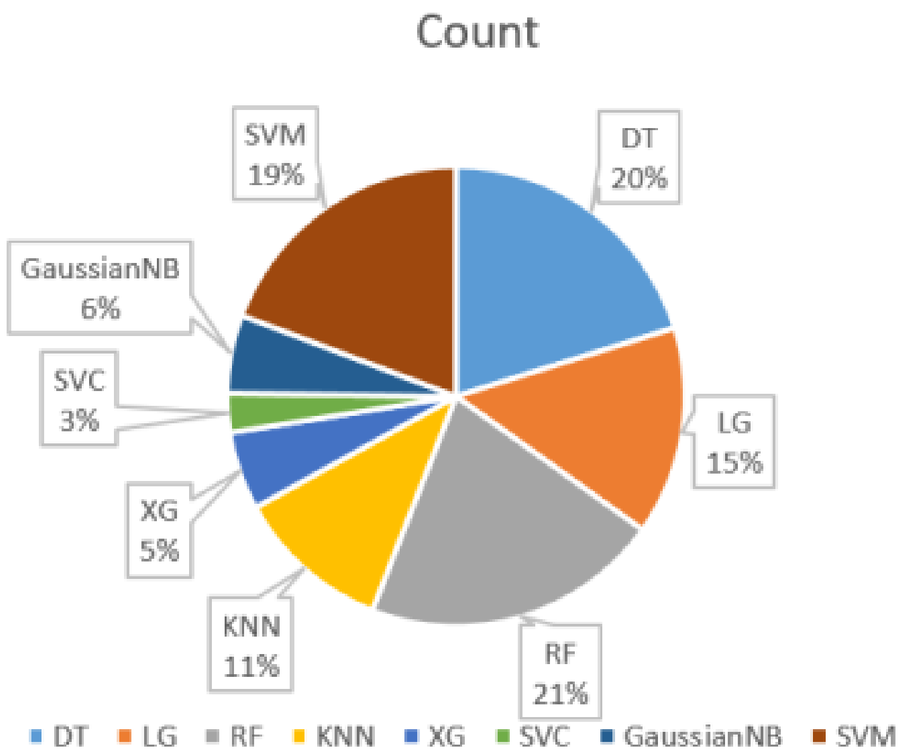
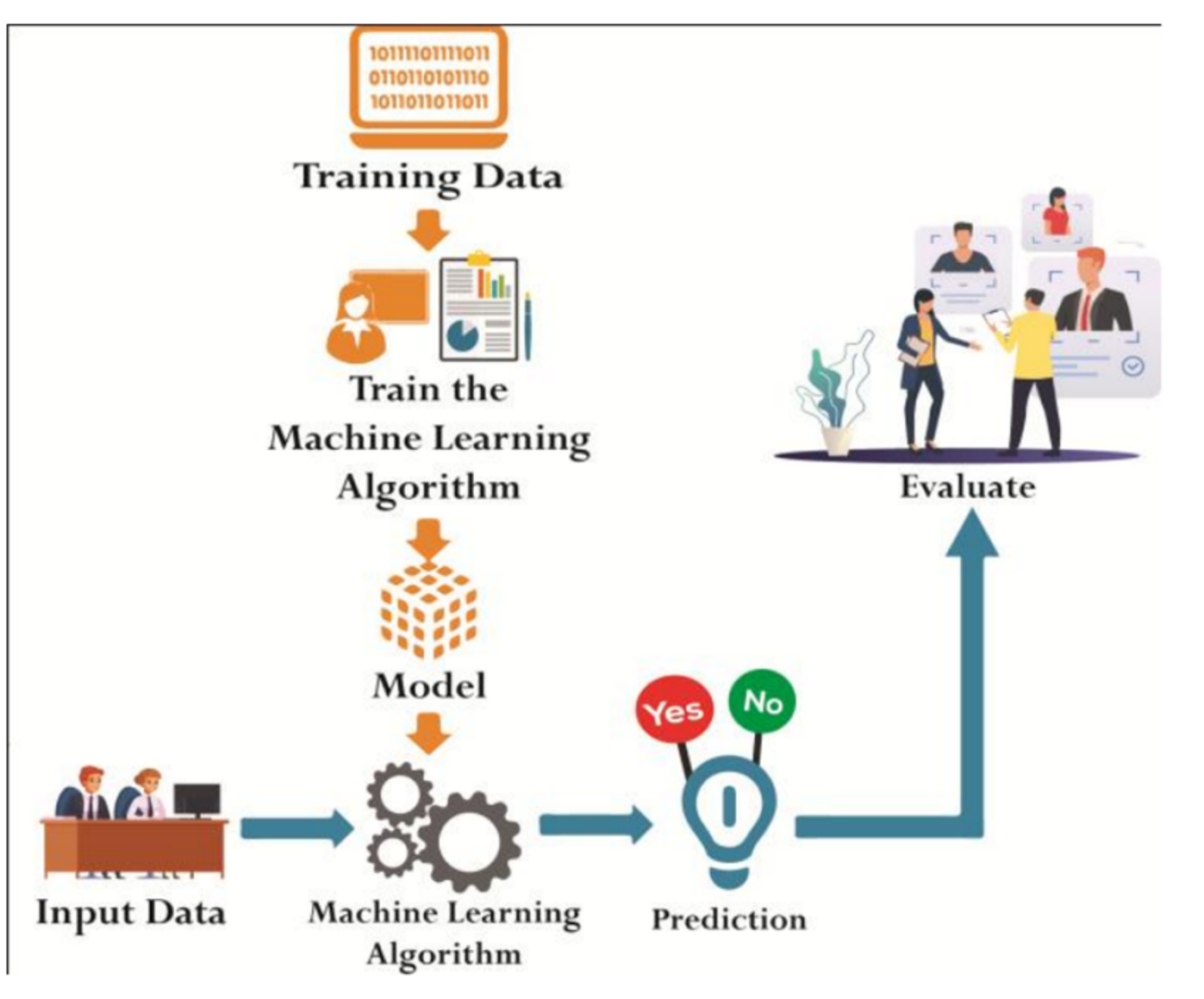


Fig 2: Number of times algorithm used

EXISTING SYSTEM



EXISTING SYSTEM

1. The dataset was pre processed and then it was trained.

2. For further processing, the dataset was split into training and testing data.

3. The machine learning model was developed.

4. After providing input to a machine learning model, it made a prediction on whether the employee will stay with the company or quit it.

5. The performance was evaluated.

DRAWBACKS OF EXISTING SYSTEM

1. As each employee has a unique job within an organisation, there is no system in place to retain individuals based on importance.

2. In the majority of articles, the IBM Dataset of Kaggle is the only dataset used, which may produce inconsistent results for other firms. Because of this, the method is not tested on real-time data.

3. In order to obtain the most beneficial features, feature selection must be done.

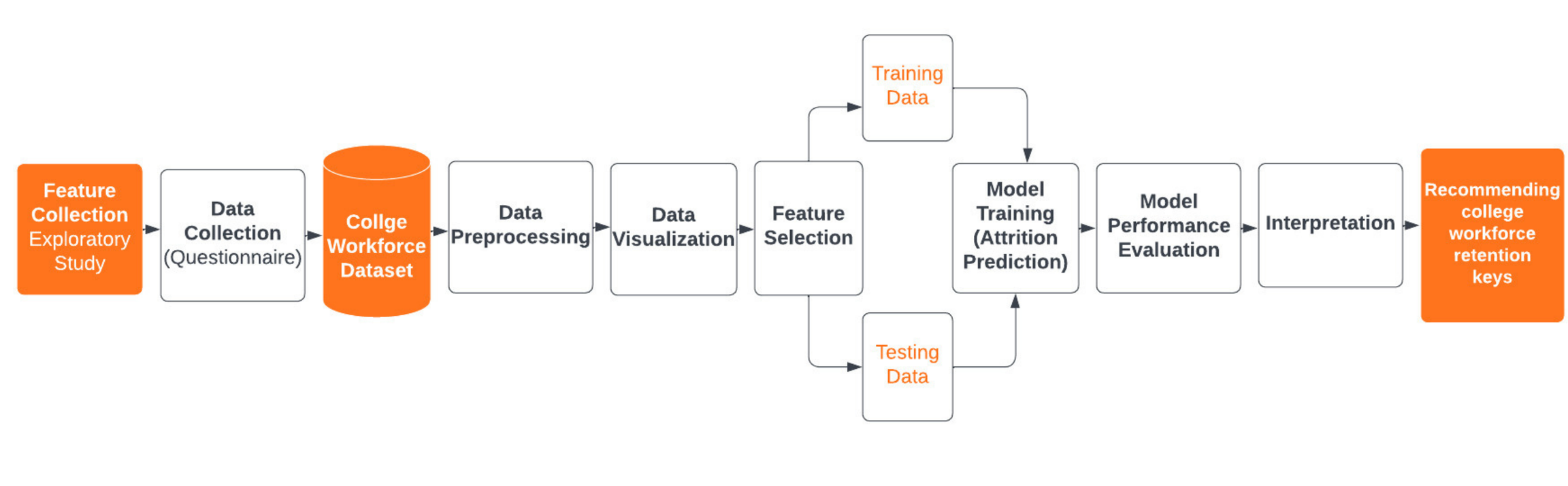
4. The current system doesn't offer recommendations for how to keep staff.

5. There are numerous systems that can forecast attrition in the IT sector, but none that can anticipate attrition among college employees.

PROPOSED SYSTEM

A web portal which will show the attrition of the oraganization based on real data collected from the college workforce by passing a questionnaire survey to the workforce. A real time dataset will be created and predictive machine learning model and ensemble learning models will be applied on the data. Best performance model will be considered for the further project. Retention of employee and data visualization are the two major features made available on the web portal.

SYSTEM ARCHITECTURE



MODULE EXPLANATION:

1. **Feature Collection**
   The first step is to identify, collect and comprehend employee features that are correct and suitable for the attrition prediction. Exploratory analysis is carried out in this step.

2. **Data Collection**
   The real data of employee is collected by an online questionnaire.

3. **Feature Selection methods**
   A feature selection process will now be used to better filter features utilizing acquired real data from our employee attrition model in order to improve our first hypothesis. Two most used methods of feature selection, Recursive Feature Elimination (a wrapper method) and SelectKBest (a filter method) will we used.

4. **Data Preprocessing**
In this step, the null value data is removed, and normalization (scaling) is performed on the data. (Data Cleaning, Normalization, Data Reduction)

5. **Feature Selection**
The factors that were chosen throughout the feature gathering steps are subjected to feature selection. (Recursive Feature Elimination and SelectKBest)

6. **Data Visualization**
Univariate and bivariate analysis are performed and hence data is visualized using python libraries using graphs and charts.

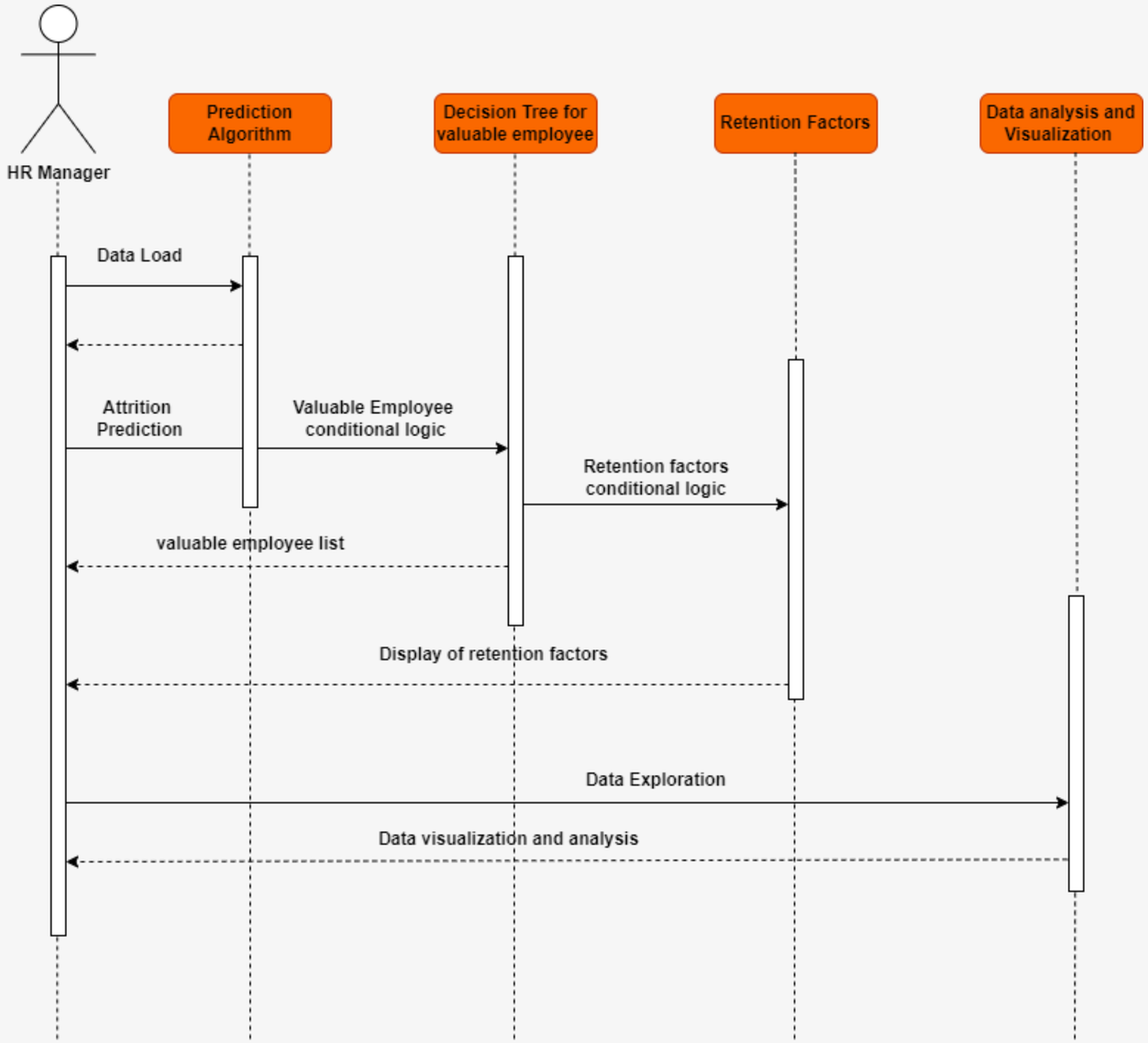
7. **Splitting**
The whole dataset is split into 10:70:20 into validation, training and test sets.

8. **Classification (LR, SVM, DT, SVM, KNN, RF, XGBoost, Adaboost, Catboost)**

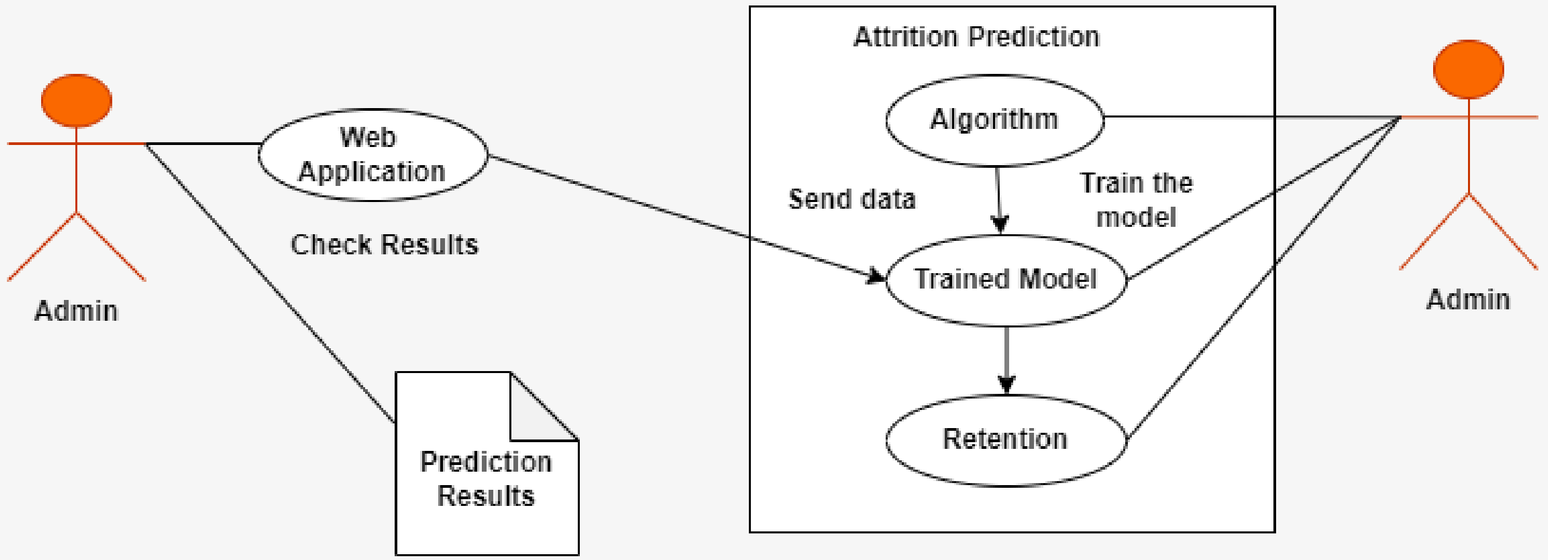
9. **Interpretation and Retention**
The process to retain the employee is performed in this step using Random forest classifier. Here job satisfaction becomes the new target variable. **(Random Forest Classifier)**

OVERCOMING THE DRAWBACKS OF EXISTING SYTEM

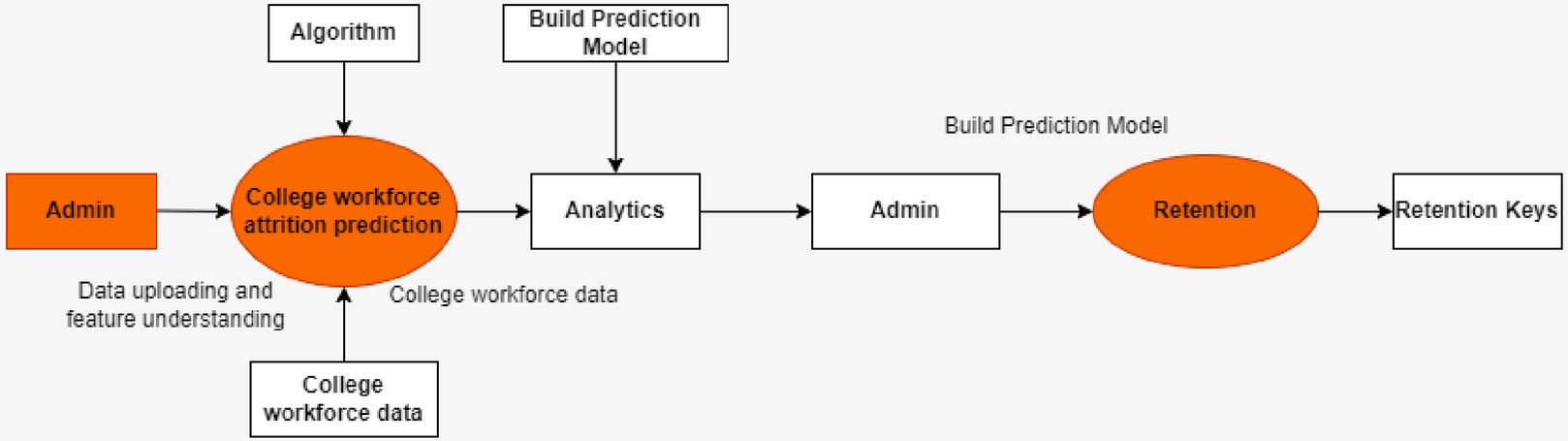
1. Creating a **dataset** of college personnel to forecast attrition outside of an IT firm.
2. Based on an **employee's relevance** to the company in relation to his or her role, attrition prediction is made.
3. As a crucial step to extracting the most crucial features, **feature selection** will be practised.
4. **Recommendations** to retain the college work force using extracted features.
5. Testing on **real time data set.**



Sequence Diagram



Use Case Diagram


Algorithm
Build Prediction Model
Admin
College workforce attrition prediction
Analytics
Admin
Build Prediction Model
Retention
Retention Keys
Data uploading and feature understanding
College workforce data
College workforce data

DFD Diagram

PROJECT PLAN

SE

1. File Copyright
2. Publish a Survey Paper
3. Learn new technologies which are required

TE

1. To collect information for real time college workforce dataset
2. Implement the project and publish the implementation paper

BE

1. Releasing and deploying the software
2. Improve it iteratively on the basis of feedback

CONCLUSION

- This understanding of attrition will aid the organisation committee in focusing on that attribute and reducing attrition by allowing them to better understand why employees leave the company.

- We looked at more than 20 papers to evaluate how other research had predicted employee turnover, and we came to the conclusion that the Random Forest method outperformed other algorithms in terms of accuracy. If implemented properly, ensemble models can potentially provide greater accuracy

REFERENCES

- Salah Al-Darraji , Dhafer G. Honi , Francesca Fallucchi , Ayad I. Abdulsada , Romeo Giuliano , Husam A. Abdulmalik, 1 Employee Attrition Prediction Using Deep Neural Networks, MDPI, 2021
- Dr. M. Subhashini1 and Dr. R. Gopinath2, 2 Employee Attrition Prediction in Industry Using Machine Learning Techniques, IJARET (Scopus Index), 2020
- Priyanka Sadana, Divya Munnuru, 3 Machine Learning Model to Predict Work Force Attrition, International Conference for Convergence in Technology (I2CT), 2021
- Praphula Kumar Jain, Madhur Jain, Rajendra Pamula, 4 Explaining and predicting employee's attrition: a machine learning approach, Springer, 2020
- Apurva Mhatre, Avantika Mahalingam, Mahadevan Narayanan, Akash Nair, Suyash Jaju, 6 Predicting Employee Attrition along with Identifying High Risk Employees using Big Data and Machine Learning, IEEE, 2020

Thank you